LA-UR-22-21078

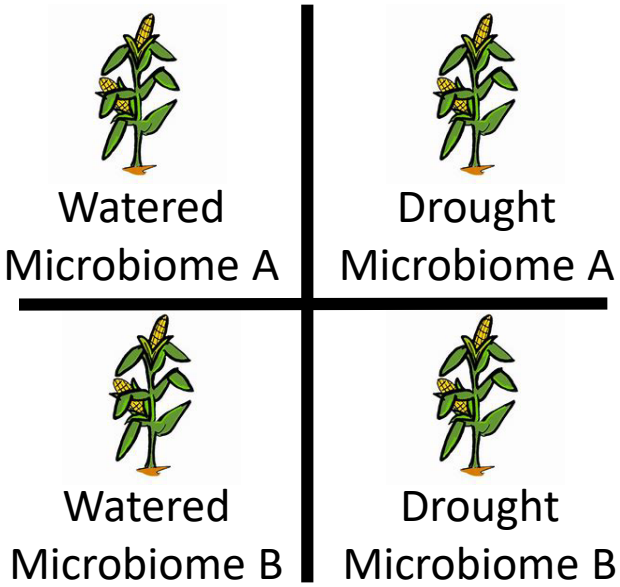
Approved for public release; distribution is unlimited.

Title:	Leveraging 2-D GC-MS and Machine Learning to Explore the Corn Metabolome
Author(s):	Mitchell, Joshua Merritt
Intended for:	I'm interviewing for a job at PathAI and I want to include a few slides about what I've been doing at LANL during my PostDoc
Issued:	2022-02-08

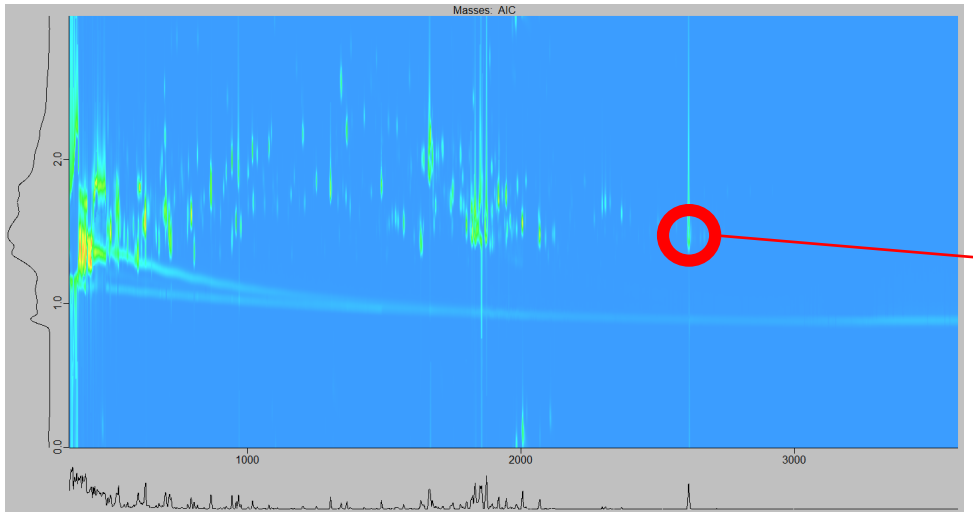
Leveraging 2-D GC-MS and Machine Learning to Explore the Corn Metabolome

Interview Presentation by Joshua Mitchell

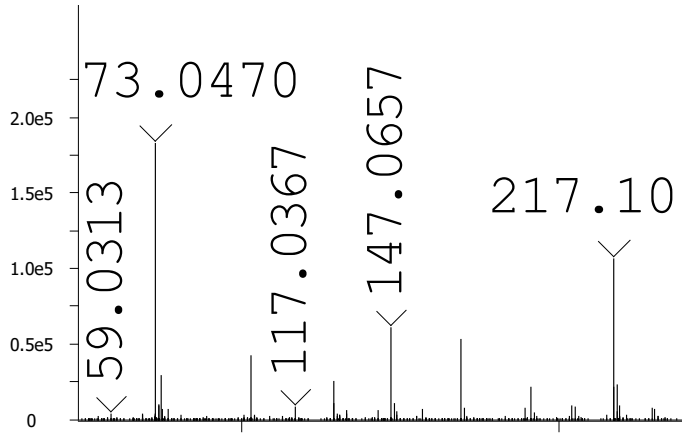
2-Dimensional Gas Chromatography/Mass Spectrometry for Corn Metabolomics



Corn Plants



2D-GC MS



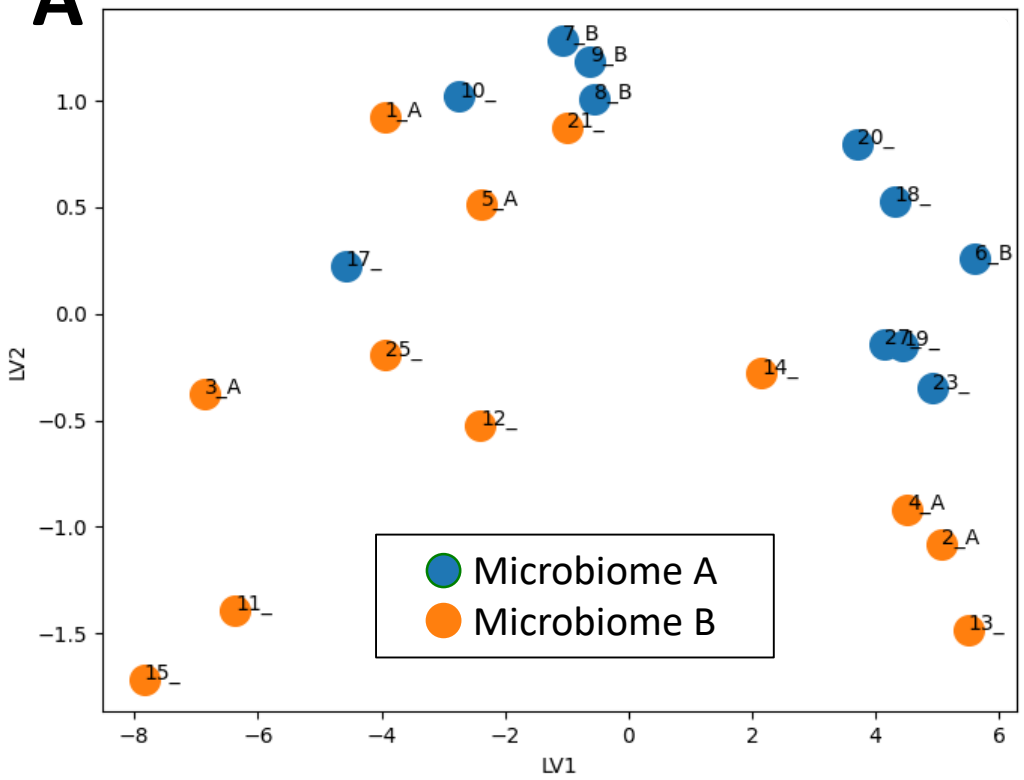
EI-MS Spectrum

Can we find a metabolic response to Watering or Microbiome?

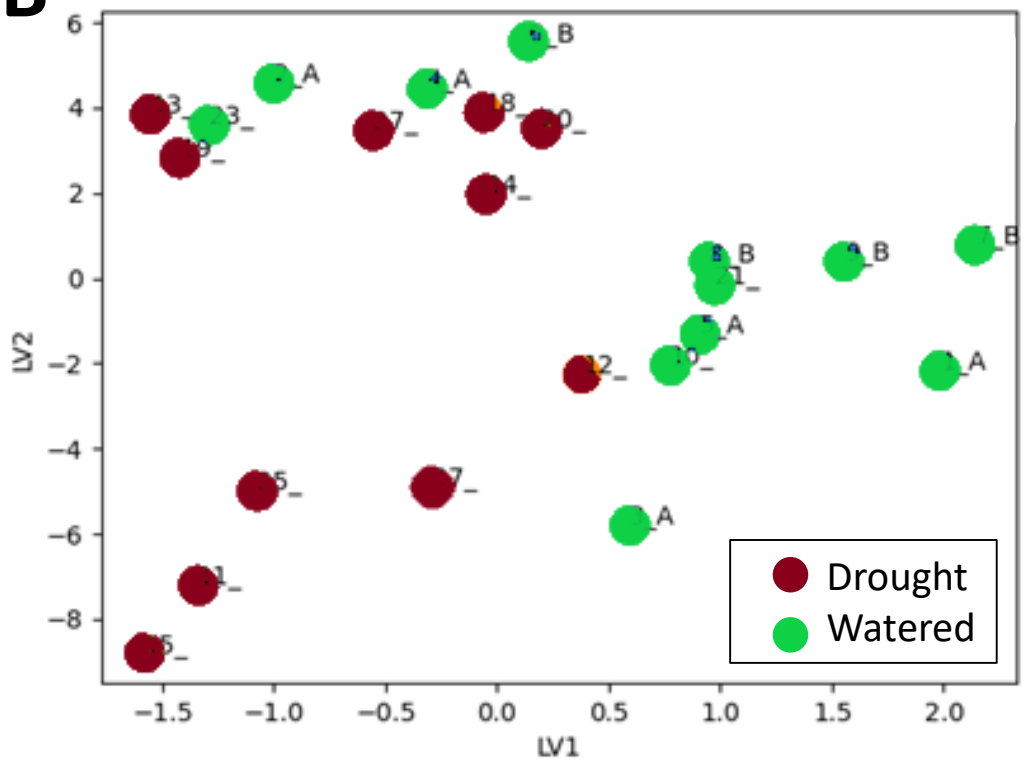
For unknown EI-MS spectra can we predict what *kind* of compound it represents? (e.g., fatty acid, amino acid, etc.)

Metabolite Profiles Respond to Drought and Microbiome Treatments

Microbiome A vs Microbiome B (Corn Roots)



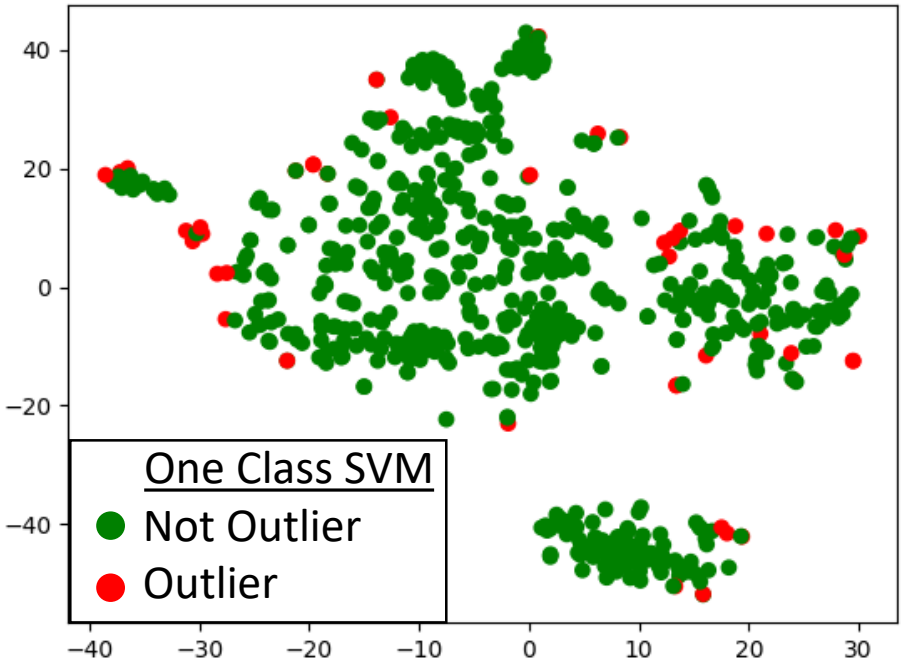
Drought vs. Watered (Corn Roots)



- PLSDA performed using normalized intensities for identified metabolites.
- Distinct response in root metabolite profiles to both microbiome and watering treatments.
- Morphology effect (large tap root vs. side root) seen in both cases (LV1 in A, LV2 in B)

We can Predict Compound Classes… (sort of)

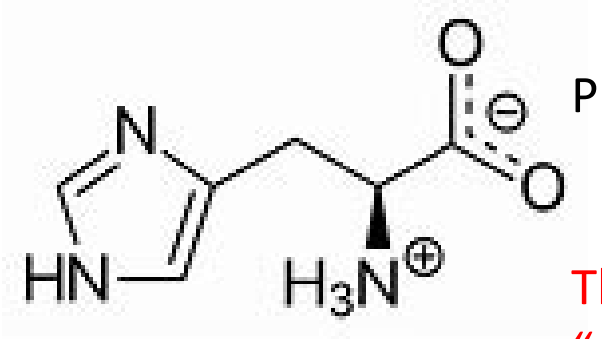
Clustering of EI-MS Spectra
t-SNE + One Class SVM (nu=0.05)



- t-SNE clusters are chemically similar
- Outliers tend to have spectra with anomalously high numbers of peaks.

Compound Class Prediction
(features constructed using pooling and NMF)

	Top-K Accuracy (2-fold stratified CV) for SVM (rbf kernel)						
	C=0.001	C=0.01	C=0.1	C=1	C=10	C=100	C=1000
k=1	0.598	0.613	0.608	0.653	0.668	0.651	0.643
k=2	0.788	0.813	0.802	0.813	0.826	0.821	0.804
k=3	0.873	0.868	0.867	0.871	0.883	0.895	0.878
k=4	0.916	0.913	0.911	0.923	0.931	0.934	0.92
k=5	0.933	0.946	0.939	0.947	0.959	0.971	0.962



"Correct" Label:
Organoheterocyclic Compound
Predicted Label:
Organic Acid Derivative

The fundamental problem is that the "correct" labels are at best incomplete.

Work performed with N.Lubbers and A.Kim

LANL Acknowledgements

Lab Directed Research and Development Team

- Dr. Sanna Sevanto (Earth and Environmental Sciences)
- Dr. Kelsey Carter (Earth and Environmental Sciences)
- Dr. Nicholas Lubbers (Information Sciences)
- Dr. Anastasiia Kim (Information Sciences)
- Dr. Chris Yeager (Chemical Diagnostics and Engineering)
- Dr. Abby Nachtsheim (Statistical Sciences)
- Jack Heneghan (Earth and Environmental Sciences)

Other

- Dr. Thomas Yoshida (Chemical Diagnostics and Engineering)

Funding Provided by LDRD